

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
2 August 2001 (02.08.2001)

PCT

(10) International Publication Number
WO 01/55911 A1

(51) International Patent Classification: **G06F 17/30**

(74) Agents: **ROBERTS, Jon, L.** et al.; Roberts Abokhair and Mardula, LLC, Suite 1000, 11800 Sunrise Valley Drive, Reston, VA 20191 (US).

(21) International Application Number: **PCT/US01/02527**

(22) International Filing Date: 26 January 2001 (26.01.2001)

(81) Designated States (*national*): DE, GB, JP.

(25) Filing Language: English

(84) Designated States (*regional*): European patent (AT, BF, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

(26) Publication Language: English

(30) Priority Data:
09/491,920 27 January 2000 (27.01.2000) US

Published:

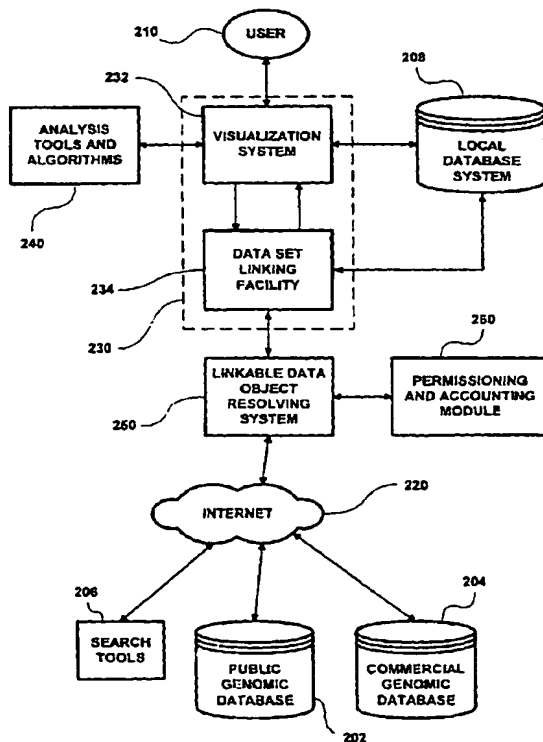
with international search report
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(71) Applicant: **INFORMAX, INC.** [US/US]; 6110 Executive Boulevard, 10th Floor, North Bethesda, MD 20852 (US).

(72) Inventor: **BABENKO, Vadim**; Informax, Inc., 6110 Executive Boulevard, 10th Floor, North Bethesda, MD (US).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: **INTEGRATED ACCESS TO BIOMEDICAL RESOURCES**



(57) Abstract: A system and method of performing genomic research via online analysis of genomic data objects. A novel technical information model and implementation facility combines genomic visualization and analysis tools with the ability to link genomic data objects across (and within) plural databases. This provides for interactivity that is not possible using conventional Internet access processes and also provides for data integration that further enhances the effectiveness of genomic research. A business process controls how the databases are accessed, particularly the commercial fee-for-access databases, so that the researcher using a system according to the invention need not be concerned with keeping track of access limitations.

WO 01/55911 A1

1 Title: **INTEGRATED ACCESS TO BIOMEDICAL RESOURCES**

2
3 **BACKGROUND OF THE INVENTION**

4 **1. Field of the Invention**

5 The present invention relates generally to the field of data processing. More
6 particularly, the present invention relates to the use of the Internet to access a variety of
7 biomedical data resources in an integrated fashion, in order to provide for increased
8 efficiency in genomic research.

9 **2. Background Information**

10 Almost every cell of every living organism contains a complete set of instructions
11 for creating that organism and regulating its cellular structures and activities over its
12 lifetime. That set of instructions is called a genome.

13 A genome is organized into distinct, microscopic units called chromosomes.
14 Chromosomes are coiled threads of deoxyribonucleic acid (DNA). Each thread of DNA is
15 composed of two long chains of nucleotides bound together in pairs to form a double
16 helix. The human genome is made up of three and a half billion of these nucleotide pairs.

17 A given DNA strand contains the cells' instructions for producing proteins. These
18 instructions are in the form of specific sequences of nucleotide bases, called genes, within
19 the DNA strand. Scientists estimate that 80,000 to 100,000 of these basic units of heredity
20 exist within the human genome. Proteins perform a wide variety of physiological tasks.
21 They facilitate processes such as digestion, breathing, immune responses, the production
22 of heat and energy, and the movement of fluids in and out of cells.

23 Most members of a species have the same collection of genes. However, each
24 individual's unique characteristics stem from slight variations in the sequence of the

1 nucleotides that comprise the genes of that individual. These slight genetic variations that
2 define unique characteristics of individuals are called polymorphisms. On average, the
3 DNA of any two individuals in a species will differ by about 0.1%.

4 Another class of variations -- called mutations -- also occurs. Both polymorphic
5 and mutagenic variations may be harmful to an individual by inhibiting the production, or
6 altering the normal function, of a protein. Most diseases result from these types of genetic
7 variations.

8 Genomics is the study of the nucleic sequences within a genome. The goal of
9 genomic inquiry is to identify the sequence of nucleotides, understand the function of
10 every gene they comprise, and clarify the genetic variations that define individuality and
11 create disease.

12 Genomics has a broad scope of applications. These range from the most basic of
13 research endeavors to the promise of diagnostic usefulness.

14 An important factor limiting the development of new drugs is the limited number
15 of known target molecules for which new drugs can be developed. Disease target
16 molecules are those that can be affected by a drug and cause a subsequent, desired
17 biological reaction in the body. Historically, the process of discovering new target
18 molecules has been extremely slow and very expensive due to reliance on trial-and-error
19 approaches to discovery. Genomic research will reduce the reliance on trial-and-error by
20 enabling drug designers to go directly to target molecules of interest. Thus, applying
21 genomic research to drug development should produce new and better drugs more quickly,
22 and at a reduced cost.

23 Another way that genomic research can help the pharmaceutical industry is in the
24 emerging field of pharmacogenomics. Pharmacogenomics focuses on identifying genetic

1 variation among patients that may affect the efficacy of drug treatment – how well an
2 individual's body absorbs and metabolizes a specific drug -- in order to develop more
3 personalized drug therapies. Nearly all drug companies are developing
4 pharmacogenomics units as a reaction to increasing evidence that a given drug does not
5 have the same effect on all people.

6 In particular, pharmacogenomics is believed to offer at least three different useful
7 applications:

- 8 • Increasing the success rate of clinical trials by improving the process of patient
9 population selection;
- 10 • Identifying new uses for existing drugs; and
- 11 • Rescuing drugs that have failed previous drug trials by identifying more
12 appropriate populations for using the drug. Candidates for drugs to be rescued
13 include those that produce adverse reactions in particular sub-populations.

14 Molecular toxicology is another area of technology that can benefit from genomics
15 research. Approximately 2.2 million Americans are admitted to hospitals every year as a
16 result of adverse side effects from drugs. Over 100,000 Americans die annually from
17 these adverse (and often unpredictable) effects. For instance, some cause liver damage,
18 while others are harmful to the kidneys. Organ-specific gene expression profiles for drugs
19 already available will enable researchers to study the toxicity of new drug compounds with
20 more certainty.

21 In addition, gene expression data, combined with polymorphism information
22 related to metabolic pathways, will provide important indications of the way an individual
23 patient will react to drugs of various dosage levels, thereby significantly reducing the
24 unwanted side effects of therapy.

1 Risk assessment is a major area of diagnostics that will benefit from genomics.
2 Historically, prediction of whether someone is at special risk for a particular disease has
3 focused on measuring general indicators in the body, such as blood pressure and
4 cholesterol levels. These measurements reflect general physiology but do not explain the
5 specific genetic basis of disease in an individual patient. Consequently, these diagnostic
6 tests do not discern the underlying cause of disease and can result in compromised medical
7 care for patients and increased risk of litigation.

8 New genomic-based diagnostics will focus on determining an individual's risk of
9 developing a particular disease by looking at specific genes and any disease-related
10 changes in that patient. These new diagnostics will likely lead to far better preventive care
11 by offering more accurate assessments of a patient's potential risk for developing a
12 particular disease.

13 Personalized medicine is another major area of diagnostics that will benefit from
14 genomics. Genomic information will be available to develop molecular diagnostic tests to
15 identify the genetic make-up of individuals. These diagnostic tests will revolutionize
16 medicine by enabling physicians to establish therapies designed for each patient, i.e.,
17 personalized medicine.

18 For example, many types of cancer that are distinct at the cellular level
19 nevertheless have similar symptoms. Because symptoms may be similar in one genetic
20 type of cancer and another, it is important to know everything possible about cancer genes
21 and their interactions in prescribing an effective treatment.

22 As another example, physicians will be able to use a molecular/genomic test to
23 help select the most effective drug with the minimum number of side effects. As a result,

1 this approach should benefit the patient with more customized care, reduced length of
2 illness, and, ultimately, a better and longer life.

3 Besides healthcare, the field of agriculture is also likely to benefit from genomic
4 research. The ability to diagnose plant and animal diseases and develop treatments
5 targeted against those diseases should produce better agricultural products and improve
6 yields. For example, the comparison of genetic information from disease-or pest-resistant
7 plant strains with non-resistant strains and the use of selective breeding programs for
8 favorable traits will significantly increase the number and success of new strains available
9 to various agricultural areas around the world. This has major implications for not only
10 increasing the quantity of food but also its nutritional quality.

11 Other fields that will likely derive important benefits from genomic information
12 include forensics, veterinary medicine, textile production, waste control, and
13 environmental remediation.

14 A significant impediment to achieving any of the foregoing expected beneficial
15 results of genomic research is the sheer size of the amount of genomic information to be
16 sifted through and studied. Without exaggeration, the amount of raw nucleic sequencing
17 data available to be sifted through and studied is unimaginably vast. Moreover, the body
18 of data gets bigger every day (literally), since newly sequenced strands of DNA are
19 documented on an ongoing basis. Because of the vastness of the genomic data, it is stored,
20 handled, and manipulated via computers.

21 Bioinformatics is the use of computers to retrieve, process, and analyze biological
22 information. This field of data processing is now considered essential for drug discovery
23 and development. Scientists are augmenting traditional "wet" biology with quantitative
24 analyses, database comparisons, and computational algorithms. In this way, biology

1 research, at least preliminarily, is conducted in a virtual environment before the scientist
2 sets foot in the laboratory. Bioinformatic tools and services assist pharmaceutical and
3 biotechnology researchers with all phases of drug discovery and development including
4 gene discovery, understanding disease pathways, identifying new disease targets and the
5 discovery and correlation of gene sequence variation to disease.

6 Unfortunately, conventional means of access to genomic information does not
7 provide for comprehensive and easy access so that the data can be analyzed or studied in a
8 computationally transparent manner. Genomic information is accumulated in a wide
9 variety of databases, some public (freely available), some commercial (available for a
10 price), and some proprietary ("in-house" resources that are not shared with outsiders).

11 Referring to **Fig. 1**, genomic databases, both public **102** and private **104**, as well as
12 search tools **106** are available online. A user **110** uses an interface device **112** to access
13 the databases **102**, **104** and search tools **106**. The interface device **112** communicates with
14 a data access portal site **120** via an Internet connection **130**. The portal **120** makes
15 connections to the databases **102**, **104** and search tools **106** via the Internet **140**. Another
16 operational mode for the user **110** to access the genomic research resources **102**, **104**, **106**
17 is serially via an Internet connection without using the portal **120**. (This simplified mode
18 of connection is not illustrated.) The interface device **112** is typically an implementation
19 of a thin client browser application.

20 Each database is made up of data objects that contain biological data. The data
21 objects differ from one database to the next in terms of the types of biological data they
22 contain, and in terms of their formats. Thus, studying data from plural databases requires
23 the researcher to learn how to interpret the data as presented in the unique data structures
24 and data contents of each particular database. This is a significant inconvenience.

1 Thus, what is needed is a bioinformatics visualization tool that will automatically
2 interpret data from diverse genomic databases (each containing genomic data objects of
3 varying formats) so that it is presented to a user in a predictable, easily-recognizable
4 format.

5 Additionally, because of the disparate storage formats used by the different
6 databases, many of the data objects to be analyzed using automated analysis tools must be
7 converted from their native format into a format that is recognizable to the automated
8 analysis tool to be used. This is a further inconvenience.

9 Thus, what is needed is a bioinformatics software tool that will automatically
10 translate data from diverse genomic databases (each containing genomic data objects of
11 varying formats) so that it is presented to analysis facilities according to a uniform format.

12 Furthermore, when it is discovered that a pair of data objects has a significant
13 relationship to one another, there is no conventional mechanism (other than manually
14 scribbling a note to oneself) for establishing a linking relationship between them. Such a
15 mechanism is most conspicuously absent in the case where the pair of data objects are
16 found in two entirely different databases.

17 Thus, what is needed is a software facility that enables a user to establish linking
18 relationships between data objects, even when those data objects are drawn from diverse
19 databases and have diverse formats and types of content.

20 Moreover, the existing way for accessing and navigating Hyper Text Markup
21 Language (HTML) genomic data files provided by Internet database hosts or by a
22 centralized Internet portal has crucial limitations. For one thing, there is no interactivity
23 possible for manipulation and/or analysis of the data. This interactivity is crucial in terms
24 of research effectiveness.

1 Thus, what is needed is a software facility that enables a user to interactively
2 access and navigate genomic data files over the Internet, whether those files be in the form
3 of HTML (as is typically done) or in other formats the data may appear.

4 SUMMARY OF THE INVENTION

5 It is an object of the present invention to provide a bioinformatics visualization tool
6 that will automatically interpret data from diverse genomic databases (each containing
7 genomic data objects of varying formats) so that it is presented to a user in a predictable,
8 easily-recognizable format.

9 It is another object of the present invention to provide a bioinformatics software
10 process and system that will automatically translate data from diverse genomic databases
11 (each containing genomic data objects of varying formats) so that it is presented to
12 analysis facilities according to a uniform format.

13 It is yet another object of the present invention to provide a software process and
14 system that enables a user to establish linking relationships between data objects, even
15 when those data objects are drawn from diverse databases and have diverse formats and
16 types of content.

17 It is still another object of the present invention to provide a software process and
18 system that enables a user to interactively access and navigate HTML (or other format)
19 genomic data files over the Internet.

20 It is a further object of the present invention to provide a business process for
21 enabling seamless access and interactivity with plural genomic databases, even in the case
22 where one or more of those databases is a commercial (i.e., fee for access) database.

Some of the above objects are made possible by a data processing system that is in electronic communication with a local genomic database system and with one or more

1 remote genomic database systems. The data processing system includes a graphical user
2 interface that enables a user to view genomic data objects graphically. It also includes a
3 genomic data object linker and a linkable data object resolver that resolves one or more
4 genomic data objects, which are linkable with respect to a subject genomic data object,
5 from among data objects found in the local genomic database system and the one or more
6 remote genomic database systems. A resolved genomic data object that is resolved by the
7 linkable data object resolver is linked to the subject genomic data object by the data object
8 linker, so that the resolved genomic data object and the subject genomic data object are
9 each provided to the graphical user interface.

10 Some of the above objects are also made possible by a system for performing
11 genomic research that is in electronic communication with a local genomic database
12 system and with one or more remote genomic database systems. The system includes a
13 means for presenting a user with a graphical view of genomic data objects, as well as a
14 means for linking genomic data objects to one another. The system further includes a
15 means for resolving a genomic data object with respect to a subject genomic data object,
16 from among genomic data objects found in the local genomic database system and the one
17 or more remote genomic database systems. A resolved genomic data object that is
18 resolved by the means for resolving is linked to the subject genomic data object by the
19 means for linking, so that the resolved genomic data object and the subject genomic data
20 object are each provided to the means for printing.

21 Another way that some of the above objects are made possible is by a method of
22 performing genomic research with respect to a subject genomic data object, using a local
23 genomic database and one or more remote genomic databases as resources. The method
24 includes the act of resolving a linkable genomic data object, with respect to the subject

1 genomic data object, from among the local genomic database and the one or more remote
2 genomic databases, regardless of the data formats of the genomic data objects.
3 Additionally, the method includes the act of linking the subject genomic data object with
4 the linkable genomic data object to form a set of linked genomic data objects.
5 Furthermore, the method includes the act of storing the set of linked genomic data objects
6 in the local genomic database.

7 Still another way that some of the above objects are made possible is by a
8 computer system for use in genomic research that implements a method as described
9 above.

10 One of the above objects is made possible by a novel technical information model
11 and implementation facility. This is accomplished by combining genomic visualization
12 and analysis tools with the ability to link genomic data objects across (and within) plural
13 databases.

14 Another of the above objects is made possible by a method of administering access
15 to a plurality of genomic databases, where the genomic databases include a local genomic
16 database, a public genomic database, and a commercial genomic database. The method
17 includes a step of resolving linkable data objects with respect to a subject data object, and
18 a further step of linking the resolved data objects to the subject genomic data object. The
19 linkable data objects that are resolved from public genomic databases are resolved
20 regardless of the data formats of genomic data objects stored therein, and without
21 restriction as to access costs. The linkable data objects that are resolved from commercial
22 genomic databases are resolved regardless of data formats of genomic data objects stored
23 therein, and are resolved subject to applicable, predetermined access agreements for the
24 commercial genomic databases.

1 Still another way that some of the above objects are made possible is by a
2 computer program product for enabling a computer to administer access to a plurality of
3 genomic databases. The plurality of genomic databases includes a local genomic
4 database, a public genomic database, and a commercial genomic database. The computer
5 program product is software instructions for enabling the computer to perform
6 predetermined operations, and a computer readable medium embodying the software
7 instructions. The predetermined operations including the acts according to the methods
8 discussed above.

9 One aspect of the present invention is a local database for storing genomic data,
10 such as nucleic acid sequences, amino acid sequences, oligonucleotides, results of Basic
11 Local Aligned Search Tool (BLAST) searches, and entries from medical databases such as
12 MEDLINE.

13 Another aspect of the present invention is a visualization and analysis facility.
14 Visualization and analysis is provide, preferably via dialog boxes, so that parameters may
15 be set for BLAST searches; so that text-based searching may be made of a sequence
16 database, and so that bibliographic search may be done of a database. A linker is included
17 for linking together nucleic acid sequences, amino acid sequences, BLAST search results,
18 MEDLINE entries, etc. All types of genomic/biomedical information are visualized via a
19 graphical interface that incorporates viewer and editor components.

20 Still another aspect of the present invention is an Internet connector having a
21 programming module that resolves links between database objects. This connector looks
22 either in the local database or in a remote Internet server to obtain a data object being
23 sought.

BRIEF DESCRIPTION OF THE DRAWING

Additional objects and advantages of the present invention will be apparent in the following detailed description read in conjunction with the accompanying drawing figures.

Fig. 1 illustrates a conventional configuration of using a browser to access Internet database resources via a portal.

Fig. 2 illustrates integrated access to Internet database resources according to an embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

One way to view the present invention is that the traditional approach of using a thin client browser and a thin portal (refer to Fig. 1) to provide a researcher with access to genomic research resources is abandoned. In its place, applicants have discovered the increased effectiveness of a system that provides an interactive interface with the resources. This interactivity is crucial for increasing the effectiveness of genomic research. Another way to view the present invention is as a technical information model and implementation facility. The interactivity aspect of the invention is provided by the combination of linking of data objects and its visualization and analysis aspects.

In addition to a local database for storing intermediate and finalized results, the present invention has a visualization and analysis aspect. Visualization is provide in an advanced form that shows the sequences and other molecules in graphical presentations that are intuitively appealing to the human perception. In addition to interactivity, the ability for the user to easily integrate (i.e., link) data and store the integrated data in the local database is very helpful.

An additional functionality that is provided in a preferred embodiment of the present invention are full-scale analysis tools and algorithms directly available at the user

1 interface rather than remotely over the Internet. Full-scale analysis allows DNA to be
2 evaluated in view of, protein, enzyme, and oligos data sets, BLAST results, MEDLINE
3 Entrez data, and amino acids.

4 An example of a software product that can provide the visualization and analysis
5 aspects of the present invention is the VectorNTI™ product of InforMax, Inc. of North
6 Bethesda, Maryland.

7 Another aspect according to a preferred embodiment of the present invention is the
8 use of a resolving system that reaches out to plural databases over the Internet (e.g., NCBI,
9 Entrez, PubMed, SRS) to provide integrated database searches.

10 Referring to Fig. 2, databases, both public 202 and private 204, and various
11 research tools 206 are available online. Also available as a research resource are the
12 previous research results and other proprietary data that user 210 stores in a local database
13 system 208. As in the prior art, the Internet 220 is used as a communication medium to
14 access the various remote resources 202, 204, 206. However, in contrast with the prior art,
15 an entirely different set of tools is used for conducting research.

16 The user 210 utilizes a user interface 230 that includes a visualization system 232
17 and a data set linking facility 234 (hereinafter "linker" for short). The linker 234 provides
18 for integration of data sets that the user deems to be worthy of being associated with one
19 another for further study in relation to one another. Data sets so linked may be more
20 closely examined or analyzed using analysis tools and algorithms 240. Examples of useful
21 analysis tools and algorithms to include for use with the present invention are BioPlot™,
22 AlignX™, and ContigExpress™, which are all products of InforMax, Inc. of North
23 Bethesda, MD. A number of other available analysis tools, such as BLAST may also be
24 usefully employed.

1 This examination and analysis produces results that may themselves be linked to
2 the data sets from which they were derived. The local database **208** is used to store the
3 integrated data sets and results for later study by the user **210**, or his or her colleagues.
4 The later study may be in the form of additional computer analysis, or in a biology
5 laboratory if the results are deemed to be sufficiently promising.

6 Candidates for linking are identified by the linkable data object resolving system
7 **250**. The linkable data object resolving system **250** connects via the Internet **220** to access
8 any of various search tools **206** and databases **202**, **204** to search for data objects that are
9 relevant to a subject data object that the user **210** has identified as being of interest. The
10 resolving system **250** does not establish links. Rather, the resolving system **250** identifies
11 data objects from the enormously vast collections of data that are available for inspection
12 over the Internet that should have a reasonable probability of being relevant to the subject
13 data object.

14 To limit and guide the searching by the resolving system **250**, the permissioning
15 and accounting module **260** directs the resolving system **250** to access only databases that
16 are public **202** or those commercial databases **204** for which access agreements have been
17 established. The accounting aspect of the permissioning and accounting module **260**
18 keeps records of access times, durations, and authorizations regarding the use of the
19 commercial databasses **204**.

20 In addition to the apparatus aspects of the present invention, methods also form
21 some aspects of the invention. A method of performing genomic research according to the
22 present invention is accomplished, first, by resolving a linkable genomic data object, and
23 then, by linking the linkable genomic data object with a subject genomic data object to
24 form a set of linked genomic data objects. The resolving act is performed with respect to

1 the subject genomic data object, using one or more remote genomic databases as
2 resources. Optionally, a local genomic database is also used in resolving a linkable
3 genomic data object. The act of resolving is performed regardless of the data formats of
4 the genomic data objects as they may be found in the various databases. After the
5 genomic data objects are linked, they are preferably stored as a set of linked genomic data
6 objects in the local genomic database.

7 Another aspect of the present invention is that it represents a business process
8 wherein predetermined access agreements for commercial databases are used to guide
9 research steps so that accessing of these databases is entirely seamless from the point of
10 view of the researcher/user employing the process. This results in a process of
11 administering access to a plurality of genomic databases, public, commercial, as well as
12 local (possibly proprietary). The process includes a step of resolving linkable data objects
13 with respect to a subject data object, and a further step of linking the resolved data objects
14 to the subject genomic data object. The linkable data objects that are resolved from public
15 genomic databases are resolved regardless of the data formats of genomic data objects
16 stored therein, and without restriction as to access costs. The linkable data objects that are
17 resolved from commercial genomic databases are resolved regardless of data formats of
18 genomic data objects stored therein, and are resolved subject to applicable, predetermined
19 access agreements for the commercial genomic databases.

20 The present invention has been described in terms of preferred embodiments,
21 however, it will be appreciated that various modifications and improvements may be made
22 to the described embodiments without departing from the scope of the invention. The
23 scope of the invention is limited only by the appended claims.

WHAT IS CLAIMED IS:

1 1. A data processing system that is in communication with a local database system
2 and that is in communication with one or more remote database systems, the data
3 processing system comprising:
4 a graphical user interface enabling a user to view data objects graphically;
5 a data object linker;
6 a linkable data object resolver that resolves one or more data objects, which are
7 linkable with respect to a subject data object, from among data objects found in the local
8 database system and the one or more remote database systems;
9 wherein a resolved data object resolved by the linkable data object resolver is
10 linked to the subject data object by the data object linker, so that the resolved data object
11 and the subject data object are each provided to the graphical user interface.

1 2. A data processing system that is in electronic communication with a local
2 genomic database system and that is in electronic communication with one or more remote
3 genomic database systems, the data processing system comprising:
4 a graphical user interface enabling a user to view genomic data objects graphically;
5 a genomic data object linker;
6 a linkable data object resolver that resolves one or more genomic data objects,
7 which are linkable with respect to a subject genomic data object, from among data objects
8 found in the local genomic database system and the one or more remote genomic database
9 systems;
10 wherein a resolved genomic data object resolved by the linkable data object
11 resolver is linked to the subject genomic data object by the data object linker, so that the

12 resolved genomic data object and the subject genomic data object are each provided to the
13 graphical user interface.

1 3. The data processing system of claim 2, wherein the data processing system
2 communicates with the one or more remote genomic database systems via a network.

1 4. The data processing system of claim 2, wherein the data processing system
2 communicates with the one or more remote genomic database systems via a network of
3 networks.

1 5. The data processing system of claim 2, wherein the data processing system
2 communicates with the one or more remote genomic database systems via the Internet.

1 6. The data processing system of claim 2, wherein the resolved genomic data
2 object and the subject genomic data object are each of a data type selected from the group
3 consisting of: nucleic acid sequences, amino acid sequences, oligonucleotides, results of a
4 BLAST search, and medical data.

1 7. The data processing system of claim 6, wherein the resolved genomic data
2 object and the subject genomic data object are of different data types.

1 8. The data processing system of claim 2, wherein the local genomic database
2 system and the one or more remote genomic database systems each contain data objects of
3 types that are selected from the group consisting of: nucleic acid sequences, amino acid
4 sequences, oligonucleotides, results of BLAST searches, and medical data.

1 9. The data processing system of claim 2, wherein the graphical user interface has
2 the capability to graphically depict nucleic acid sequences, amino acid sequences,
3 oligonucleotides, results of BLAST searches, and medical data.

1 10. The data processing system of claim 2, wherein the genomic data object linker
2 links genomic data objects that are of differing data types.

1 11. A system for performing genomic research, the system being in electronic
2 communication with a local genomic database system and that is in electronic
3 communication with one or more remote genomic database systems, the system
4 comprising:

5 means for presenting a user with a graphical view of genomic data objects;
6 means for linking genomic data objects to one another; and
7 means for resolving a genomic data object with respect to a subject genomic data
8 object, from among genomic data objects found in the local genomic database system and
9 the one or more remote genomic database systems;

10 wherein a resolved genomic data object resolved by the means for resolving is
11 linked to the subject genomic data object by the means for linking, so that the resolved
12 genomic data object and the subject genomic data object are each provided to the means
13 for printing.

1 12. The system for performing genomic research of claim 11, wherein the system
2 communicates with the one or more remote genomic database systems via the Internet.

1 13. The system for performing genomic research of claim 11, wherein the resolved
2 genomic data object and the subject resolved genomic data object are each of a data type

3 selected from the group consisting of: nucleic acid sequences, amino acid sequences,
4 oligonucleotides, results of a BLAST search, and medical data.

1 14. The system for performing genomic research of claim 13, wherein the resolved
2 genomic data object and the subject resolved genomic data object are of different data
3 types.

1 15. The system for performing genomic research of claim 11, wherein the local
2 genomic database system and the one or more remote genomic database systems each
3 contain data objects of types that are selected from the group consisting of: nucleic acid
4 sequences, amino acid sequences, oligonucleotides, results of BLAST searches, and
5 medical data.

1 16. The system for performing genomic research of claim 11, wherein the means
2 for presenting has the capability to graphically depict nucleic acid sequences, amino acid
3 sequences, oligonucleotides, results of BLAST searches, and medical data.

1 17. The system for performing genomic research of claim 11, wherein the means
2 for linking links genomic data objects that are of differing data types.

1 18. A computer system adapted to genomic research with respect to a subject
2 genomic data object, using a local genomic database and remote genomic databases as
3 resources, the computer system comprising:

4 a processor, and

5 a memory, in electronic communication with the processor, including software
6 instructions adapted to enable the computer system to perform the steps of:

7 resolve a linkable genomic data object, with respect to the subject genomic data
8 object, from among the local genomic database and the remote genomic
9 databases, regardless of the data formats of the genomic data objects;
10 link the subject genomic data object with the linkable genomic data object to
11 form a set of linked genomic data objects; and
12 store the set of linked genomic data objects in the local genomic database.

1 19. A method of performing genomic research with respect to a subject genomic
2 data object, using a local genomic database and one or more remote genomic databases as
3 resources, the method comprising:

4 resolving a linkable genomic data object, with respect to the subject genomic data
5 object, from among the local genomic database and the one or more remote genomic
6 databases, regardless of the data formats of the genomic data objects;
7 linking the subject genomic data object with the linkable genomic data object to
8 form a set of linked genomic data objects; and
9 storing the set of linked genomic data objects in the local genomic database.

1 20. The method of performing genomic research of claim 19, wherein the resolved
2 genomic data object and the subject genomic data object are each of a data type selected
3 from the group consisting of: nucleic acid sequences, amino acid sequences,
4 oligonucleotides, results of a BLAST search, and medical data.

1 21. The method of performing genomic research of claim 20, wherein the resolved
2 genomic data object and the subject genomic data object are of different data types.

1 22. The method of performing genomic research of claim 19, wherein the local
2 genomic database system and the one or more remote genomic database systems each
3 contain data objects of types that are selected from the group consisting of: nucleic acid
4 sequences, amino acid sequences, oligonucleotides, results of BLAST searches, and
5 medical data.

1 23. The method of performing genomic research of claim 19, the method further
2 comprising:
3 providing a graphical user interface to graphically depict the subject genomic data
4 object and the resolved genomic data object, regardless of whether they are nucleic acid
5 sequences, amino acid sequences, oligonucleotides, results of BLAST searches, or medical
6 data.

1 24. The method of performing genomic research of claim 19, wherein the act of
2 linking links genomic data objects that are of differing data types.

1 25. A method of administering access to a plurality of genomic databases, the
2 plurality of genomic databases including a local genomic database, a public genomic
3 database, and a commercial genomic database, the method comprising:
4 resolving one or more linkable data objects with respect to a subject data object;
5 and
6 linking the one or more resolved data objects to said subject genomic data object;
7 wherein the one or more linkable data objects are resolved from public genomic
8 databases, regardless of data formats of genomic data objects stored therein, and without
9 restriction as to access costs; and

10 wherein the one or more linkable data objects are resolved from commercial
11 genomic databases, regardless of data formats of genomic data objects stored therein, and
12 subject to predetermined access agreements for the commercial genomic databases.

1 26. The method of administering access to a plurality of genomic databases recited
2 in claim 25, wherein the resolved genomic data object and the subject genomic data object
3 are each of a data type selected from the group consisting of: nucleic acid sequences,
4 amino acid sequences, oligonucleotides, results of a BLAST search, and medical data.

1 27. The method of administering access to a plurality of genomic databases recited
2 in claim 26, wherein the resolved genomic data object and the subject genomic data object
3 are of different data types.

1 28. The method of administering access to a plurality of genomic databases recited
2 in claim 25, wherein the local genomic database system and the one or more remote
3 genomic database systems each contain data objects of types that are selected from the
4 group consisting of: nucleic acid sequences, amino acid sequences, oligonucleotides,
5 results of BLAST searches, and medical data.

1 29. The method of administering access to a plurality of genomic databases recited
2 in claim 25, the method further comprising:

3 providing a graphical user interface to graphically depict the subject genomic data
4 object and the resolved genomic data object, regardless of whether they are nucleic acid
5 sequences, amino acid sequences, oligonucleotides, results of BLAST searches, or medical
6 data.

1 30. The method of administering access to a plurality of genomic databases recited
2 in claim 25, wherein the act of linking links genomic data objects that are of differing data
3 types.

1 31. A computer program product for enabling a computer to perform genomic
2 research with respect to a subject genomic data object, using a local genomic database and
3 remote genomic databases as resources, the computer program product comprising:
4 software instructions for enabling the computer to perform predetermined
5 operations, and
6 a computer readable medium embodying the software instructions;
7 the predetermined operations including the acts of:
8 resolve a linkable genomic data object, with respect to the subject genomic data
9 object, from among the local genomic database and the remote genomic
10 databases, regardless of the data formats of the genomic data objects;
11 link the subject genomic data object with the linkable genomic data object to
12 form a set of linked genomic data objects; and
13 store the set of linked genomic data objects in the local genomic database.

1 32. A computer program product for enabling a computer to administer access to a
2 plurality of genomic databases, the plurality of genomic databases including a local
3 genomic database, a public genomic database, and a commercial genomic database, the
4 computer program product comprising:
5 software instructions for enabling the computer to perform predetermined
6 operations, and
7 a computer readable medium embodying the software instructions;

8 the predetermined operations including the acts of:
9 resolve one or more linkable data objects with respect to a subject data object;
10 and
11 linking the one or more resolved data objects to said subject genomic data
12 object;
13 wherein the one or more linkable data objects are resolved from public
14 genomic databases, regardless of data formats of genomic data objects
15 stored therein, and without restriction as to access costs; and
16 wherein the one or more linkable data objects are resolved from commercial
17 genomic databases, regardless of data formats of genomic data objects
18 stored therein, and subject to predetermined access agreements for the
19 commercial genomic databases.

1/2

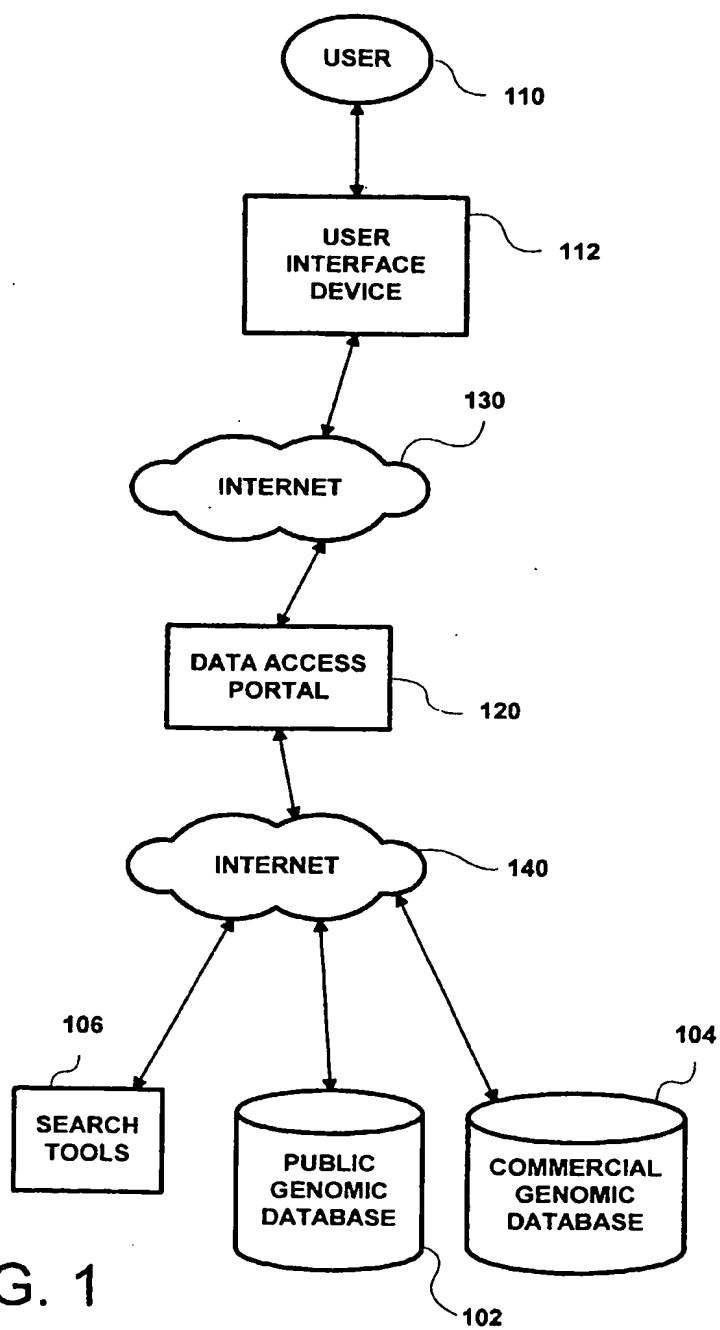


FIG. 1
PRIOR ART

2/2

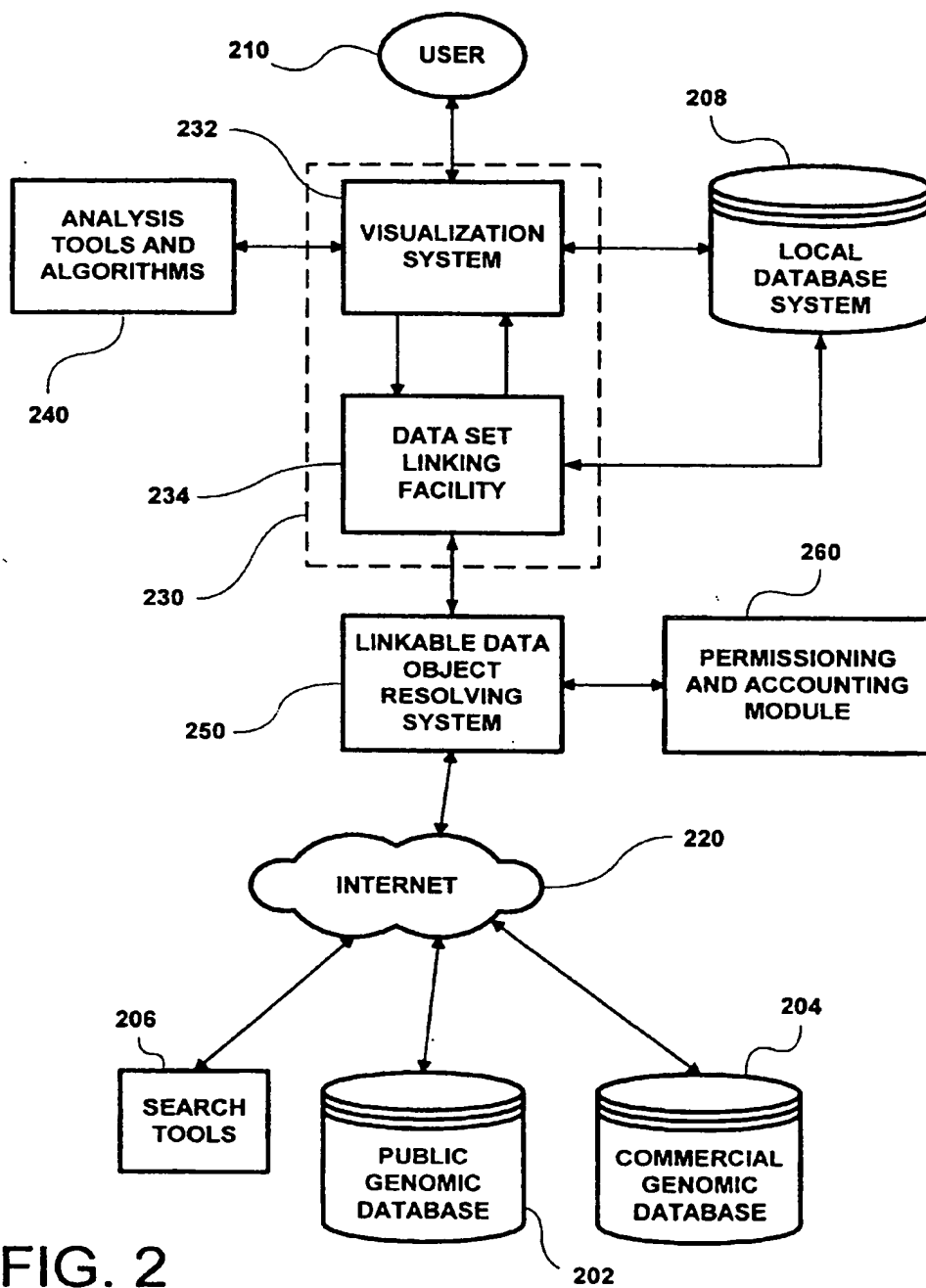


FIG. 2

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/02527

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, EPO-Internal, PAJ, INSPEC, IBM-TDB

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DAVIDSON S.B. ET AL.: "BioKleisli: a digital library for biomedical researchers" INTERNATIONAL JOURNAL ON DIGITAL LIBRARIES, vol. 1, no. 1, April 1997 (1997-04), pages 36-53, XP002167355 Germany paragraph '04.4! - paragraph '0005!; figure 6 abstract --- -/--	1-32

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *8* document member of the same patent family

Date of the actual completion of the international search

15 May 2001

Date of mailing of the international search report

29/05/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Deane, E

INTERNATIONAL SEARCH REPORT

Internat- Application No
PCT/US 01/02527

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>ACHARD F. : "Automatic generation of links between heterogeneous genomic databases"</p> <p>PROC. FIRST INT. SYMPOSIUM ON INTELLIGENCE IN NEURAL AND BIOLOGICAL SYSTEMS, 29 - 31 May 1995, pages 78-83, XP002167356</p> <p>Herndon VA, USA</p> <p>abstract</p> <p>page 81, left-hand column, line 1 -page 82, right-hand column, line 12</p> <p>---</p>	<p>1,2,11, 18,19, 25,31,32</p>
A	<p>YEE D.P. ET AL.: "Automated clustering and assembly of large EST collections"</p> <p>PROC. 6TH. INT. CONF. ON INTELLIGENT SYSTEMS FOR MOELCULAR BIOLOGY, 28 June 1998 (1998-06-28) - 1 July 1998 (1998-07-01), pages 203-211, XP000994591</p> <p>Montreal, Canada</p> <p>abstract</p> <p>---</p>	<p>1,2,11, 18,19, 25,31,32</p>
A	<p>MOUSHENG XU ET AL.: "Associated Biological Information Retrieval From Distributed Databases"</p> <p>PROC. 1998 ACM CIKM 7TH. INT. CONF. ON INFORMATION AND KNOWLEDGE MANAGEMENT, 3 - 7 November 1998, pages 193-200, XP002167357</p> <p>Bethesda, MD, USA</p> <p>page 193, left-hand column, line 1 -page 195, left-hand column, line 31</p> <p>-----</p>	<p>1,2,11, 18,19, 25,31,32</p>